

Speech-Based Emotion Recognition Using Deep Learning

Summary:

Human emotions play a crucial role in communication, and recognizing emotions from speech can improve human-computer interactions. This project aims to develop a deep learning model that can analyze voice patterns and classify emotions. The goal is to enhance applications in healthcare, virtual assistants, and customer service by making AI more emotionally intelligent.

Objectives:

- Developing a deep learning model for classifying emotions from speech.
- Using advanced speech processing techniques to extract meaningful features.
- Improving the accuracy of emotion recognition using state-of-the-art models.
- Comparing different deep learning models to determine the best-performing approach.

Methodology:

- **Dataset:** Use publicly available datasets like RAVDESS, CREMA-D, IEMOCAP, and TESS.
- **Feature Extraction:** Utilizing Mel-Frequency Cepstral Coefficients (MFCC), spectrograms, and pitch analysis to preprocess audio signals.
- **Model Selection:**
 - Train CNNs for spectrogram-based classification.
 - Use LSTMs/RNNs for sequential speech data analysis.
 - Experimentation with transformer-based models like Wav2Vec for improved accuracy.
- **Implementation:**
 - Use **Librosa** for audio preprocessing.
 - Training models using **TensorFlow/Keras or PyTorch**.
 - Augmentation of data to handle noise variations and improve generalization.

Evaluation:

- **Qualitative Analysis:** Visualize spectrograms and observe patterns in different emotions.
- **Quantitative Analysis:**
 - Evaluation of accuracy, precision, recall, and F1-score.
 - Using confusion matrices to analyze misclassifications.
 - Comparing model performance against benchmark results from existing research.

CSE – 676B Deep Learning
Final Project Proposal
Nischith Adavala – 50591641 Rakshith Reddy Dargula – 50591661
Sumanth Yalamanchili - 50604274

- **Comparison:** We will be comparing CNNs, LSTMs, and transformer-based approaches to determine the most effective model.

Dataset:

- **RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song)** – Contains speech samples labeled with different emotions.
- **CREMA-D (Crowd-Sourced Emotional Multimodal Actors Dataset)** – High-quality dataset with diverse speech samples.
- **IEMOCAP (Interactive Emotional Dyadic Motion Capture Database)** – Includes real-world conversations with emotion labels.
- **TESS (Toronto Emotional Speech Set)** – Speech samples designed for emotion classification tasks.

References:

- Research papers on speech-based emotion recognition.
- Studies on deep learning models for audio classification.
- TensorFlow/PyTorch documentation for implementing deep learning models.
- Librosa documentation for speech feature extraction.

Github Project board Link:

<https://github.com/users/Nisch9/projects/2/views/1>